Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 09:11:10 2000; MasPar time 3.26 Seconds
167.001 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-26
Description: (1-23) from US09331631.pep
Perfect Score: 183
Sequence: 1 VKEDHQFETRGEILECYRLCQQQ 23

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 21.302; Variance 67.176; scale 0.317

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	183	100.0	23	1	W62839	Stenocarpus sinuatus a	4.06e-13
2	74	40.4	17	1	W62840	Stenocarpus sinuatus a	2.67e+00
3	71	38.8	143	1	W42635	Protein sequence that	5.52e+00
4	67	36.6	611	1	W02157	Periplasmic Beta-N-ace	1.43e+01
5	67	36.6	611	1	W85599	Hexosaminidase enzyme.	1.43e+01
6	64	35.0	373	1	R48703	G-protein coupled huma	2.89e+01
7	64	35.0	373	1	W02675	G-protein coupled huma	2.89e+01
8	64	35.0	446	1	R15498	Human dopamine D1 rece	2.89e+01
9	64	35.0	446	1	R38364	Human dopamine D1 rece	2.89e+01
10	64	35.0	446	1	R15499	Rat dopamine D1 recept	2.89e+01
11	64	35.0	487	1	W09795	D1 dopamine receptor.	2.89e+01
12	64	35.0	487	1	R13596	D1 dopamine receptor.	2.89e+01
13	64	35.0	529	1	W19001	Feline herpes virus ty	2.89e+01
14	63	34.4	768	1	W98108	Caenorhabditis elegans	3.65e+01
15	63	34.4	1608	1	R98619	Borna disease virus po	3.65e+01
16	63	34.4	1711	1	R98605	Borna disease virus po	3.65e+01
17	61	33.3	1852	1	W79897	Tumourogenic BRCA1 pro	5.77e+01
18	61	33.3	1852	1	W10011	Protein encoded by mut	5.77e+01
19	61	33.3	1863	1	R81543	BRCA1 mutant from PM24	5.77e+01
20	61	33.3	1863	1	R81485	BRCA1 mutant from pati	5.77e+01
21	61	33.3	1863	1	W76099	Human BRCA1 omi2 prote	5.77e+01
22	61	33.3	1863	1	R81524	BRCA1 mutant from samp	5.77e+01
23	61	33.3	1863	1	W10003	Protein encoded by mut	5.77e+01
24	61	33.3	1863	1	R81488	BRCA1 mutant from pati	5.77e+01
25	61	33.3	1863	1	W79889	Tumourogenic BRCA1 pro	5.77e+01
26	61	33.3	1863	1	R81529	BRCA1 mutant from samp	5.77e+01
27	61	33.3	1863	1	R81546	BRCA1 mutant from PM27	5.77e+01
28	61	33.3	1863	1	R81505	BRCA1 mutant from samp	5.77e+01
29	61	33.3	1863	1	W26522	Human BRCA1 consensus	5.77e+01
30	61	33.3	1863	1	R81522	BRCA1 mutant from samp	5.77e+01
31	61	33.3	1863	1	R81533	BRCA1 mutant from PM01	5.77e+01
32	61	33.3	1863	1	W76100	Human BRCA1 omi3 prote	5.77e+01
33	61	33.3	1863	1	R99440	BRCA1 allele #8403 tra	5.77e+01
34	61	33.3	1863	1	R81532	BRCA1 mutant from samp	5.77e+01
35	61	33.3	1863	1	R81493	BRCA1 mutant from pati	5.77e+01
36	61	33.3	1863	1	R81542	BRCA1 mutant from PM23	5.77e+01
37	61	33.3	1863	1	R97128	BRCA1, breast and ovar	5.77e+01
38	61	33.3	1863	1	R81486	BRCA1 mutant from pati	5.77e+01
39	61	33.3	1863	1	R81519	BRCA1 mutant from samp	5.77e+01
40	61	33.3	1863	1	R81534	BRCA1 mutant from PM03	5.77e+01
41	61	33.3	1863	1	R81545	BRCA1 mutant from PM26	5.77e+01
42	61	33.3	1863	1	R91208	BRCA1, breast and ovar	5.77e+01
43	61	33.3	1863	1	R81544	BRCA1 mutant from PM25	5.77e+01
44	61	33.3	1863	1	R81537	BRCA1 mutant from PM12	5.77e+01
45	61	33.3	1863	1	R81511	BRCA1 mutant from samp	5.77e+01

ALIGNMENTS

```
RESULT    1
ID   W62839 standard; Protein; 23 AA.
AC   W62839;
DT   27-OCT-1998  (first entry)
DE   Stenocarpus sinuatus antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Stenocarpus sinuatus.
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR   WPI; 98-377279/32.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 65; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   23 AA;

  Query Match          100.0%;  Score 183;  DB 1;  Length 23;
  Best Local Similarity 100.0%;  Pred. No. 4.06e-13;
  Matches   23;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 VKEDHQFETRGEILECYRLCQQQ 23
          |||||||||||||||||||||||
Qy      1 VKEDHQFETRGEILECYRLCQQQ 23


RESULT    2
ID   W62840 standard; Protein; 17 AA.
AC   W62840;
DT   27-OCT-1998  (first entry)
DE   Stenocarpus sinuatus antimicrobial protein.
KW   antimicrobial protein; infestation; control.
OS   Stenocarpus sinuatus.
FH   Key             Location/Qualifiers
FT   Misc_difference 13
FT                   /note= "undefined amino acid"
PN   WO9827805-A1.
PD   02-JUL-1998.
PF   22-DEC-1997; AU0874.
PR   20-DEC-1996; AU-004275.
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI   Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
```

```
DR   WPI; 98-377279/32.
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
PS   Claim 1; Page 66; 96pp; English.
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
SQ   Sequence   17 AA;

  Query Match           40.4%;  Score 74;  DB 1;  Length 17;
  Best Local Similarity  61.5%;  Pred. No. 2.67e+00;
  Matches     8;  Conservative    2;  Mismatches   3;  Indels   0;  Gaps   0;

Db       4 RSQILGCYLXCQQ 16
           |::|| ||  |||
Qy      10 RGEILECYRLCQQ 22


RESULT    3
ID   W42635 standard; Protein; 143 AA.
AC   W42635;
DT   22-OCT-1998  (first entry)
DE   Protein sequence that is specific for Neisseria meningitidis.
KW   N. gonorrheae; N. lactamica; chromosome Z2491; region 1; region 2;
KW   region 3; pathogenicity; blood-brain barrier; diagnosis; infection;
KW   meningitis.
OS   Neisseria meningitidis.
PN   WO9802547-A2.
PD   22-JAN-1998.
PF   11-JUL-1997; F01295.
PR   12-JUL-1996; FR-008768.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA   (SMIK ) SMITHKLINE BEECHAM.
PI   Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J,
PI   Vinals C;
DR   WPI; 98-110594/10.
DR   N-PSDB; V03553.
PT   Genes present in Neisseria meningitidis but not other Neisseria
PT   species - and related host cells, RNA, anti-sense sequences,
PT   polypeptide(s) and antibodies, useful for diagnosing Neisseria
PT   meningitidis infection and in protective vaccines
PS   Claim 8; Pages 90-91; 150pp; French.
CC   W42633-37 and W42639-41 are encoded by a DNA sequence found in region 2
CC   of Neisseria meningitidis. The specification describes DNA sequences
CC   that are found in N. meningitidis, but not in N. gonorrheae or
CC   N. lactamica, except for the genes involved in biosynthesis of the
CC   capsule polysaccharide, frpA or C, opc, porA, rotamase, sequence IC1106,
CC   IgA protease, pillin, pilC, proteins which bind transferrin and opacity
CC   proteins. The DNA sequences are found on chromosome Z2491, mainly (or
CC   within 20 kb) between tufA and pilT (region 1), pilQ and lambda-740
CC   (region 2) or argF and opaB (region 3). The DNA sequences are responsible
CC   for the differences in pathogenicity between N. meningitidis and
CC   N. gonorrheae, specifically they include the genes that allow
CC   N. meningitidis to cross the blood-brain barrier. DNA sequences common
CC   to N. meningitidis and N. gonorrheae, but absent from N. lactamica, are
CC   responsible for colonisation and penetration of the mucosa. The DNA
CC   sequences can be used to produce probes and primers, and antibodies
CC   produced against the encoded proteins are used in standard
CC   hybridisation/immunoassay processes for diagnosis of N. meningitidis
CC   infection, particularly meningitis.
SQ   Sequence   143 AA;

  Query Match           38.8%;  Score 71;  DB 1;  Length 143;
  Best Local Similarity  53.8%;  Pred. No. 5.52e+00;
  Matches     7;  Conservative    4;  Mismatches   2;  Indels   0;  Gaps   0;

Db      30 QFETKDTLLECFK 42
           ||||:  :|||::
Qy       6 QFETRGEILECYR 18


RESULT    4
ID   W02157 standard; Protein; 611 AA.
AC   W02157;
DT   14-JAN-1997  (first entry)
DE   Periplasmic Beta-N-acetylglucosaminidase.
KW   Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW   Beta-N-acetylglucosamidase; chitin; oligosaccharide; catabolic;
KW   catabolism.
OS   Vibrio furnissii.
PN   WO9625424-A1.
PD   22-AUG-1996.
PF   13-FEB-1996; U02332.
PR   13-FEB-1995; US-386727.
PA   (UYJO ) UNIV JOHNS HOPKINS.
PI   Bassler B,  Chitlaru E,  Keyhani N,  Roseman S,  Rowe C;
PI   Yu C;
DR   WPI; 96-393335/39.
DR   N-PSDB; T36388.
PT   Chitin biosynthetic enzymes end I, exo I and exo II  - are
PT   periplasmic chito:dextrinase(s), periplasmic beta-GlcNAcidase(s) and
PT   aryl beta-N-acetyl:gluco:amidase(s), respectively
PS   Claim 10; Page 73-75; 101pp; English.
CC   Periplasmic chitodextrinase (W02156), periplasmic
CC   Beta-N-acetylglucosaminidase (W02157) and aryl
CC   Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
CC   oligosaccharides with the structure (GlcNAc)n where n is 2 or
CC   higher, by contacting them with soluble chitin.  The enzymes are
CC   encoded by the genes endI, exoI and exoII respectively.  They are
CC   all genes involved in the catabolic pathway of chitin.
SQ   Sequence   611 AA;

  Query Match           36.6%;  Score 67;  DB 1;  Length 611;
  Best Local Similarity  38.9%;  Pred. No. 1.43e+01;
  Matches     7;  Conservative    5;  Mismatches   6;  Indels   0;  Gaps   0;

Db     254 IKDAPRFKYRGMMLDCAR 271
           :|:  :|  || :|:| |
Qy       1 VKEDHQFETRGEILECYR 18


RESULT    5
ID   W85599 standard; protein; 611 AA.
AC   W85599;
DT   02-MAR-1999  (first entry)
DT   08-MAR-1999  (revised)
DE   Hexosaminidase enzyme.
KW   Hexosaminidase; enzyme; laundry; cleaning agent; hydrolysis;
KW   anti-microbial; detergent; surfactant.
OS   Unidentified.
PN   WO9850512-A1.
PD   12-NOV-1998.
PF   05-MAY-1998; U09125.
PR   19-AUG-1997; US-056132.
PR   06-MAY-1997; US-045756.
PA   (PROC ) PROCTER & GAMBLE CO.
PI   Convents AC, Moese RL, Wolff AM;
DR   WPI; 99-024116/02.
DR   N-PSDB; V83129.
PT   Laundry and cleaning compositions containing hexosaminidase - to
PT   provide antimicrobial activity and remove biofilm
PS   Claim 2; Page 38-39; 64pp; English.
CC   Novel hexosaminidase enzymes (W85599-605) can be used in
CC   combination in an aqueous laundry or cleaning product. The cleaning
CC   product is used especially used to launder fabrics and to clean
CC   dishes and tableware, particularly in an automatic dishwasher, but
CC   may also be used generally as hard surface cleaner.  The cleaning
CC   product imparts antimicrobial activity and/or eliminates biofilm,
CC   the hexosaminidases having a minimum inhibitory concentration of
CC   less than about 0.125% but more preferably less than about 0.025%.
CC   (NB: entry was revised to change incorrect cross references in
CC   Comments field).
SQ   Sequence   611 AA;

  Query Match           36.6%;  Score 67;  DB 1;  Length 611;
```

```
  Best Local Similarity  38.9%;  Pred. No. 1.43e+01;
  Matches     7;  Conservative    5;  Mismatches   6;  Indels   0;  Gaps   0;

Db     254 IKDAPRFKYRGMMLDCAR 271
           :|:  :|  || :|:| |
Qy       1 VKEDHQFETRGEILECYR 18


RESULT    6
ID   R48703 standard; Protein; 373 AA.
AC   R48703;
DT   05-JUN-1996  (first entry)
DE   G-protein coupled human dopamine D1 receptor protein.
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;
KW   muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
KW   rhodopsin; opsin; odorant; cytomegalovirus.
OS   Homo sapiens.
PN   WO9405695-A1.
PD   17-MAR-1994.
PF   09-SEP-1993; U08528.
PR   10-SEP-1992; US-943236.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 94-101120/12.
PT   Polypeptides of G-coupled receptor proteins (GPRs) - useful for
PT   binding GPR ligands or modulating GPR binding
PS   Disclosure; Page 76-77; 160pp; English.
CC   Proteins R48685-R48758 represent a range of G-protein coupled receptor
CC   proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other G-protein coupled receptors.  The
CC   receptor proteins were used to design polypeptides, pref. based on the
CC   transmembrane domains, for use in G-protein coupled receptor ligand
CC   binding assays.  The polypeptide fragments retain biological activity
CC   such as binding a GPR ligand or modulating GPR ligand binding to a GPR
CC   (see R48759-R48758, R50569-R50807 and R89189-R89195 for examples of
CC   polypeptide fragments).  The polypeptide fragments can be used in
CC   compositions for treating subjects suffering from a pathology related to
CC   a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
SQ   Sequence   373 AA;

  Query Match           35.0%;  Score 64;  DB 1;  Length 373;
  Best Local Similarity  75.0%;  Pred. No. 2.89e+01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db     314 LLGCYRLC 321
           :| |||||
Qy      13 ILECYRLC 20


RESULT    7
ID   W02675 standard; peptide; 373 AA.
AC   W02675;
DT   12-NOV-1996  (first entry)
DE   G-protein coupled human dopamine D1 receptor.
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW   muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW   odorant; cytomegalovirus; serotonergic.
OS   Homo sapiens.
PN   US5508384-A.
PD   16-APR-1996.
PF   10-SEP-1992; 943236.
PR   10-SEP-1992; US-943236.
PR   09-SEP-1993; US-118270.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 96-208785/21.
PT   New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT   for treating schizophrenia
PS   Disclosure; Column 77-80; 184pp; English.
CC   Proteins W02657-W02730 represent a range of G-protein coupled receptor
CC   (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other GPR proteins.  The receptor proteins
CC   were used to design polypeptides, pref. based on the transmembrane
CC   domains, for use in G-protein coupled receptor ligand binding assays.
CC   The polypeptide fragments retain biological activity such as binding a
CC   GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
CC   for examples of polypeptide fragments).  The polypeptide fragments can
CC   be used in compositions for treating subjects suffering from a pathology
CC   related to a GPR abnormality e.g. a  psychotic disorder such as
CC   schizophrenia.
SQ   Sequence   373 AA;

  Query Match           35.0%;  Score 64;  DB 1;  Length 373;
  Best Local Similarity  75.0%;  Pred. No. 2.89e+01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db     314 LLGCYRLC 321
           :| |||||
Qy      13 ILECYRLC 20


RESULT    8
ID   R15498 standard; Protein; 446 AA.
AC   R15498;
DT   08-MAR-1992  (first entry)
DE   Human dopamine D1 receptor.
KW   catecholamine; G-protein-coupled receptor; neurotransmitter;
KW   adenylyl cyclase stimulation.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   modified_site   5
FT                   /label= OTHER
FT                   /note= "N-glycosylation site - putative"
FT   modified_site   175
FT                   /label= OTHER
FT                   /note= "N-glycosylation site - putative"
FT   domain          23. .49
FT                   /label= transmembrane
FT                   /note= "I"
FT   domain          62. .87
FT                   /label= transmembrane
FT                   /note= "II"
FT   domain          94. .120
FT                   /label= transmembrane
FT                   /note= "III"
FT   domain          138. .164
FT                   /label= transmembrane
FT                   /note= "IV"
FT   domain          192. .218
FT                   /label= transmembrane
FT                   /note= "V"
FT   domain          273. .299
FT                   /label= transmembrane
FT                   /note= "VI"
FT   domain          311. .337
FT                   /label= transmembrane
FT                   /note= "VII"
FT   region          265. .268
FT                   /label= Protein_kinase_A_phosphorylation_site
FT                   /note= "putative"
FT   region          338. .343
FT                   /label= Protein_kinase_A_phosphorylation_site
FT                   /note= "putative"
PN   WO9118005-A.
PD   28-NOV-1991.
PF   13-MAY-1991; U03308.
PR   14-MAY-1990; US-523237.
PR   17-SEP-1990; US-583852.
PA   (UYDU-) DUKE UNIV.
PA   (UYOR-) OREGON HEALTH SCI UNIV.
PI   Bunzow JR, Civelli O, Grandy DK, Zhou OY, Caron MG;
PI   Dearry A, Falardeau P, Gingrich JA;
```

```
DR   WPI; 91-369177/50.
DR   N-PSDB; Q14954.
PT   Cloned gene encoding D1-dopamine receptor - useful for e.g. drug
PT   screening, diagnosis of e.g. Parkinson's disease or schizophrenia
PT   or in gene therapy
PS   Example 1; Fig 1B; 52pp; English.
CC   The D1 dopamine receptor sequence was obtained from two overlapping
CC   clones, one genomic (HGL26) and the other from a human retina cDNA
CC   library (D233). The amino acid sequence was deduced from the
CC   nucleotide coding sequence. The receptor is similar to known
CC   G-protein coupled proteins, e.g. Cys(351) in the carboxyl terminus
CC   near transmembrane VII is conserved in most G-protein-coupled
CC   receptors; it may be palmitoylated. The carboxyl tail also contains
CC   several putative sites for phosphorylation by an agonist-dependent
CC   receptor kinase.
SQ   Sequence   446 AA;

  Query Match           35.0%;  Score 64;  DB 1;  Length 446;
  Best Local Similarity  75.0%;  Pred. No. 2.89e+01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    344 LLGCYRLC 351
          :| |||||
Qy     13 ILECYRLC 20


RESULT    9
ID   R38364 standard; Protein; 446 AA.
AC   R38364;
DT   17-AUG-1994  (first entry)
DE   Human dopamine D1 receptor.
KW   Dopamine D1 receptor; adenylate cyclase stimulation;
KW   cAMP-dependent protein kinase activation; psychomotor disorders;
KW   intronless gene; G-linked receptor family; neuron growth;
KW   neuron differentiation.
OS   Homo sapiens.
PN   CA2024096-A.
PD   28-FEB-1992.
PF   27-AUG-1990; 024096.
PR   27-AUG-1990; CA-024096.
PA   (SEEM/) SEEMAN P.
PI   Niznik HB,  Odowd BF,  Seeman P,  Sunahara R;
DR   WPI; 93-197541/25.
DR   N-PSDB; Q43964.
PT   Nucleotide base sequence of human dopamine D1 receptor -
PT   comprises a region susceptible to restriction enzyme esp. Eco R1,
PT   useful for studying associated genetic disease
PS   Disclosure; Fig 1b; 7pp; English.
CC   The full-length DNA sequence for the human dopamine D1 receptor
CC   was isolated from a lambda EMBL3 SP6-T7 human genomic library.
CC   The library was probed by a known 450bp rat D1 clone. Three
CC   positive clones with inserts of ca. 14kb were isolated. One of
CC   the clones was restriction analysed and sequenced to reveal an
CC   open reading frame of 1476 bp encoding a 446 amino acid protein.
CC   The D1 receptor belongs to the G-protein fammily; it regulates
CC   neuron growth and differentiation, influences behaviour and
CC   modifies D2 receptor-mediated events.
SQ   Sequence   446 AA;

  Query Match           35.0%;  Score 64;  DB 1;  Length 446;
  Best Local Similarity  75.0%;  Pred. No. 2.89e+01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    344 LLGCYRLC 351
          :| |||||
Qy     13 ILECYRLC 20


RESULT   10
ID   R15499 standard; Protein; 446 AA.
AC   R15499;
DT   08-MAR-1992  (first entry)
DE   Rat dopamine D1 receptor.
KW   catecholamine; G-protein-coupled receptor; neurotransmitter;
KW   adenylyl cyclase stimulation.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   modified_site   4
FT                   /label= OTHER
FT                   /note= "N-glycosylation site - putative"
FT   modified_site   174
FT                   /label= OTHER
FT                   /note= "N-glycosylation site - putative"
FT   modified_site   135
FT                   /label= Protein_kinase_A_phosphorylation_site
FT                   /note= "putative"
FT   modified_site   229
FT                   /label= Protein_kinase_A_phosphorylation_site
FT                   /note= "putative"
FT   modified_site   268
FT                   /label= Protein_kinase_A_phosphorylation_site
FT                   /note= "putative"
PN   WO9118005-A.
PD   28-NOV-1991.
PF   13-MAY-1991; U03308.
PR   14-MAY-1990; US-523237.
PR   17-SEP-1990; US-583852.
PA   (UYDU-) DUKE UNIV.
PA   (UYOR-) OREGON HEALTH SCI UNIV.
PI   Bunzow JR, Civelli O, Grandy DK, Zhou OY, Caron MG;
PI   Dearry A, Falardeau P, Gingrich JA;
DR   WPI; 91-369177/50.
DR   N-PSDB; Q14955.
PT   Cloned gene encoding D1-dopamine receptor - useful for e.g. drug
PT   screening, diagnosis of e.g. Parkinson's disease or schizophrenia
PT   or in gene therapy
PS   Example 9; Fig 3A; 52pp; English.
CC   The D1 dopamine receptor sequence was obtained from two overlapping
CC   clones, one genomic and the other from a rat striatum cDNA
CC   library. The amino acid sequence was deduced from the
CC   nucleotide coding sequence. The receptor is similar to the human D1
CC   dopamine receptor (see R15498) and known G-protein coupled proteins,
CC   e.g. Cys(351) in the carboxyl terminus is conserved in most G-protein
CC   -coupled receptors; it may be palmitoylated. The carboxyl tail also
CC   contains several putative sites for phosphorylation by an agonist-
CC   dependent receptor kinase.
SQ   Sequence   446 AA;

  Query Match           35.0%;  Score 64;  DB 1;  Length 446;
  Best Local Similarity  75.0%;  Pred. No. 2.89e+01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    344 LLGCYRLC 351
          :| |||||
Qy     13 ILECYRLC 20


RESULT   11
ID   W09795 standard; Protein; 487 AA.
AC   W09795;
DT   11-JUN-1997  (first entry)
DE   D1 dopamine receptor.
KW   D1 dopamine receptor; activation; adenylyl cyclase activity; coupled;
KW   guanine nucleotide binding regulatory protein; drug assessment; agonist;
KW   antagonist; efficacy; affinity.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   modified_site   45
FT                   /note= "N-linked glycosylation site"
PN   US5610282-A.
PD   11-MAR-1997.
PF   06-JUL-1990; 548714.
PR   06-JUL-1990; US-548714.
PR   11-MAR-1993; US-029917.
PR   19-MAY-1995; US-444734.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
```

PI Mahan LC, Mcvittie LD, Monsma FJ, Sibley DR;
DR WPI; 97-178452/16.
DR N-PSDB; T63657.
PT DNA encoding D1 dopamine receptor protein - for production of
PT transformed cells used for drug screening
PS Claim 1; Column 11-16; 24pp; English.
CC The sequence is the rat D1 dopamine receptor protein which is linked to
CC the activation of adenylyl cyclase activity. The receptor also couples
CC with guanine nucleotide binding regulatory (G) proteins. By constructing
CC cell lines that express the D1 receptor, the affinities and efficacies of
CC agonist and antagonist drugs can be assessed.
SQ Sequence 487 AA;

Query Match 35.0%; Score 64; DB 1; Length 487;
Best Local Similarity 75.0%; Pred. No. 2.89e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 385 LLGCYRLC 392
:| |||||
Qy 13 ILECYRLC 20

RESULT 12
ID R13596 standard; Protein; 487 AA.
AC R13596;
DT 01-NOV-1991 (first entry)
DE D1 dopamine receptor.
KW Adenylyl cyclase; G protein; neurotransmitter; hormone; signal;
KW transduction.
OS Rattus rattus.
PN US7548714-A.
PD 23-JUL-1991.
PF 06-JUL-1990; 154559.
PR 06-JUL-1990; US-548714.
PA (USSH) NAT INST OF HEALTH.
PI Sibley D, Monsma F, Mahan L, Mcvitte L;
DR WPI; 91-260183/35.
DR N-PSDB; Q13337.
PT DNA encoding D1, dopamine receptor - used to investigate
PT affinities and efficacy of agonist and antagonist drugs with the
PT D1 receptor.
PS Disclosure; Fig 1A; 38pp; English.
CC The sequence was deduced from a clone isolated from a rat striatal
CC cDNA library. The gene can be ligated into expression vectors
CC for prodn. of the D1 dopamine receptor. Hydrophobicity analysis
CC revealed seven transmembrane spanning domains. The NH2 terminus
CC contains one N-glycosylation site and the third cytoplasmic loop
CC exhibits one recognition site for phosphorylation by the cAMP-
CC dependent protein kinase. The long COOH terminus contains several
CC serine and threonine residues possibly representing additional
CC sites for regulatory phosphorylation. The receptor couples with
CC guanine binding regulatory protein (G protein) and is linked to
CC the stimulation of adenylyl cyclase. D1 receptor expressimg cell
CC lines can be used to investigate the affinities and efficacies of
CC agonist and antagonist drugs. For diagnostic purposes, expression
CC of the receptor can be measured using e.g. antibodies to the
CC receptor.
SQ Sequence 487 AA;

Query Match 35.0%; Score 64; DB 1; Length 487;
Best Local Similarity 75.0%; Pred. No. 2.89e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 385 LLGCYRLC 392
:| |||||
Qy 13 ILECYRLC 20

RESULT 13
ID W19001 standard; Protein; 529 AA.
AC W19001;
DT 05-MAY-1998 (first entry)
DE Feline herpes virus type 1 truncated ORF1-encoded protein.
KW Feline herpes virus type 1; open reading frame; ORF; vector; vaccine;
KW feline infectious peritonitis virus; FIPV; cat; immunisation;
KW rhinotracheitis.
OS Feline herpesvirus.
PN WO9720059-A1.
PD 05-JUN-1997.
PF 19-NOV-1996; F01830.
PR 30-NOV-1995; FR-014450.
PA (INMR) RHONE MERIEUX SA.
PI Audonnet JCF, Baudu PGN, Riviere MAE;
DR WPI; 97-310613/28.
DR N-PSDB; T69857.
PT Live recombinant vaccine based on feline herpes virus - has
PT antigen-encoding sequence inserted in open reading frame 2 or 5,
PT particularly for protection against feline infectious peritonitis
PT virus
PS Claim 18; Fig 1; 60pp; French.
CC This sequence represent part of protein encoded by ORF1 from the feline
CC herpes virus type 1 (FHV-1) from strain CO. The nucleotide sequence
CC is used as a vector to generate a live recombinant vaccine, in which a
CC polypeptide coding sequence (especially a gene taken from the feline
CC infectious peritonitis virus (FIPV)) is inserted into open reading
CC frames 5 and/or 2. Vaccines comprising the vector are used to protect
CC cats, specifically against FIPV. The vaccine is attenuated but retains
CC a good capacity to replicate in vivo and still protects against
CC infectious rhinotracheitis (caused by FHV).
SQ Sequence 529 AA;

Query Match 35.0%; Score 64; DB 1; Length 529;
Best Local Similarity 41.2%; Pred. No. 2.89e+01;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 165 HFDTRGDALTTSNLCTE 181
:|:|||: | || :
Qy 6 QFETRGEILECYRLCQQ 22

RESULT 14
ID W98108 standard; Protein; 768 AA.
AC W98108;
DT 21-JUN-1999 (first entry)
DE Caenorhabditis elegans elongation factor-2 kinase (eEF-2 kinase).
KW Elongation factor-2 kinase; eEF-2 kinase; nematode; protein kinase;
KW inhibitor; breast cancer; therapy.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT Region 66. .79
FT /note= "predicted amphipathic alpha-helix
FT structure"
PN WO9909199-A2.
PD 25-FEB-1999.
PF 20-AUG-1998; U17272.
PR 20-AUG-1997; US-914999.
PA (UYNE-) UNIV NEW JERSEY.
PI Hait WN, Pavur KS, Ryazanov AG;
DR WPI; 99-181050/15.
DR N-PSDB; X24907.
PT New isolated protein kinase, eEF-2 - used to develop agents for
PT controlling the amount, or activity of protein kinases, e.g. for
PT treating cancers or other hyperproliferative pathologies
PS Claim 6; Page 128-130; 195pp; English.
CC This protein is Caenorhabditis elegans elongation factor-2 kinase
CC (eEF-2 kinase), a member of a new superfamily of eukaryotic protein
CC kinases that phosphorylate within an alpha-helical domain of a
CC target protein, as opposed to beta-turns as seen in all other
CC protein kinases. eEF-2 kinase is a ubiquitous enzyme involved in
CC the regulation of protein synthesis and the cell cycle. It has no
CC homology to any other mammalian protein kinase, and is therefore an
CC ideal target in the search for a specific protein kinase inhibitor.
CC Since preliminary evidence suggests that human eEF-2 kinase (see
CC W98106) is upregulated in human cancers, including breast cancer,
CC identification of specific inhibitors of eEF-2 kinase may lead to
CC the development of novel anticancer drugs. Assays have been

```
CC   developed utilising eEF-2 kinase and a phosphorylation target
CC   (see W98109) to facilitate high-throughput screening for compounds
CC   that can specifically inhibit eEF-2 kinase.  Methods of assessing
CC   eEF-2 kinase levels for diagnostic purposes, and therapeutic
CC   formulations to inhibit eEF-2 kinase activity are also disclosed.
CC   Sequences complementary to eEF-2 kinase may have therapeutic
CC   efficacy as antisense drugs or be used in gene therapy.  A
CC   ribozyme that cleaves eEF-2 kinase mRNA is also claimed.
SQ   Sequence   768 AA;

  Query Match           34.4%;  Score 63;  DB 1;  Length 768;
  Best Local Similarity  63.6%;  Pred. No. 3.65e+01;
  Matches     7;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    125 ARGAMRECYRL 135
          :|| : |||||
Qy      9 TRGEILECYRL 19


RESULT   15
ID   R98619 standard; Protein; 1608 AA.
AC   R98619;
DT   10-DEC-1996  (first entry)
DE   Borna disease virus polymerase.
KW   Borna disease virus; BDV; G-protein; p57; nervous system disease;
KW   neuro-psychiatric disease; schizophrenia; diagnosis; therapy;
KW   vaccine; antibody.
OS   Borna disease virus strain V.
PN   WO9621020-A2.
PD   11-JUL-1996.
PF   05-JAN-1996; U00418.
PR   06-JAN-1995; US-369822.
PR   04-MAY-1995; US-434831.
PR   04-JAN-1996; US-582776.
PA   (REGC ) UNIV CALIFORNIA.
PI   Briese T,  Kliche S,  Lipkin WI,  Schneemann A,  Schneider PA;
PI   Stitz L;
DR   WPI; 96-333995/33.
DR   N-PSDB; T38104.
PT   Borna disease virus (BDV) nucleotide and protein sequences - useful
PT   for the diagnosis and treatment of infection and non-BDV related
PT   neuro-logic and neuro-psychiatric disease
PS   Claim 2; Fig 2; 186pp; English.
CC   Borna disease virus (BDV) polymerase (R98619), or pol or p180, was
CC   identified from an ORF on the virus antigenome strand (T38104).
CC   The amino acid sequence for pol after splice modification is given
CC   in R98605.  Recombinant pol (recpol) can be expressed in transformed
CC   host (partic. mammalian) cells.  It is useful in assays for
CC   detecting BDV infection and for diagnosing non-BDV related neurologic
CC   and neuropsychiatric diseases.  It may also be incorporated into
CC   vaccines and used to raise anti-BDV antibodies.
SQ   Sequence   1608 AA;

  Query Match           34.4%;  Score 63;  DB 1;  Length 1608;
  Best Local Similarity  42.1%;  Pred. No. 3.65e+01;
  Matches     8;  Conservative    3;  Mismatches   8;  Indels   0;  Gaps   0;

Db   1324 VEETNDFTARGHHHGCYSL 1342
          | | ::| :||    || |
Qy      1 VKEDHQFETRGEILECYRL 19


Search completed: Sat May 13 09:11:17 2000
Job time : 7 secs.
```